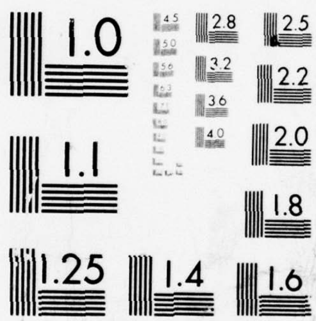
AD-A072 513 CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER F/G 12/1
A RANDOM GRAPH.(U)
APR 79 S M ROSS N00014-77-C-0299
UNCLASSIFIED ORC-79-5 NL

1 OF 1
AD
A072513

END
DATE
FILMED
9-79
DDC



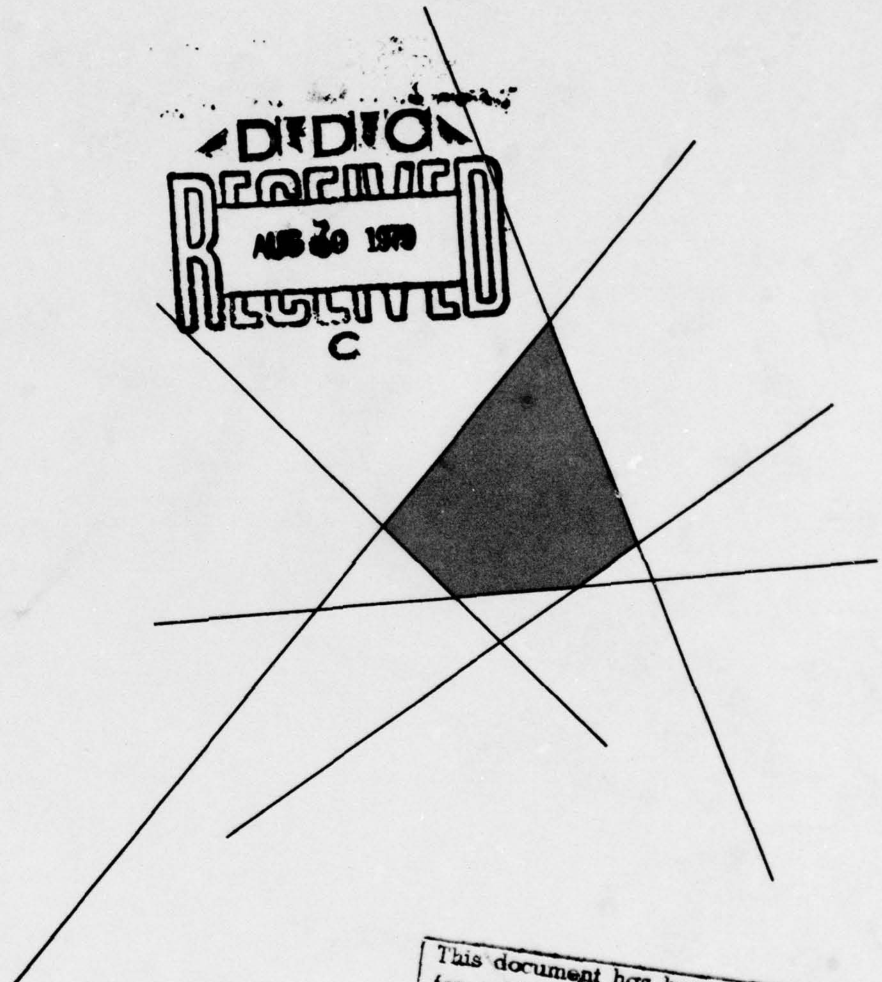
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ORC 79-5
APRIL 1979

LEVEL

A RANDOM GRAPH

by
SHELDON M. ROSS



OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

A RANDOM GRAPH

by

Sheldon M. Ross
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

APRIL 1979

ORC 79-5

This research has been partially supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3213B and the Office of Naval Research under Contract N00014-77-C-0299 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 79-5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) A RANDOM GRAPH		5. TYPE OF REPORT & PERIOD COVERED (9) Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Sheldon M. Ross		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-77-C-0299 AFOSR-77-3213
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720	61102F	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 2304/A5 (17) A5
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE (11) April 1979
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 16p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(14) ORC-79-5

18. SUPPLEMENTARY NOTES

Also supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3213B.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random Graph
Connected
Schur Convex

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

79 08 8 015

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE 270 750 Unclassified
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Let $X(1), X(2), \ldots, X(n)$ be independent random variables such that

sum from j=1 to n of

$$P\{X(i) = j\} = P_j \ , \ j = 1,2, \ldots, n \ , \ \sum_{1}^{n} P_j = 1$$

sub j sub j

and consider a graph with n nodes numbered $1,2, \ldots, n$ and the arcs $(i, X(i))$, $i = 1,2, \ldots, n$. We determine the probability that the above so-called random graph is connected and then develop a recursive formula for the distribution of C , the number of connected components it contains. We also derive expressions for the mean and variance of C .

1473B

A RANDOM GRAPH

by

Sheldon M. Ross

0. INTRODUCTION AND SUMMARY

Let $X(1), X(2), \ldots, X(n)$ be independent random variables such that

$$P\{X(i) = j\} = P_j \ , \ j = 1,2, \ldots, n \ , \ \sum_1^n P_j = 1$$

and consider a graph with n nodes numbered $1,2, \ldots, n$ and the arcs $(i,X(i))$, $i = 1,2, \ldots, n$. In Section 1 we determine the probability that the above so-called random graph is connected and then develop a recursive formula for the distribution of C , the number of connected components it contains. We also derive expressions for the mean and variance of C .

The above problem has previously been considered in [1], [3], [4], [5] by a different approach than the one we employ and only for the special case $P_j \equiv 1/n$. However even in this special case our formula for $\text{Var}(C)$ appears to be new. In the final section we show by Schur convexity arguments that the probability of a connected graph is minimized and $E(C)$ is maximized when $P_j \equiv 1/n$.

1. MAIN RESULTS

Before obtaining the desired probability that the graph is connected we shall consider a related problem having $r + 1$ nodes - $0, 1, \ldots, r$ - and r arcs $(i, Y(i))$, $i = 1, 2, \ldots, r$, where the Y_i are independent and such that $P\{Y_i = j\} = Q_j$, $j = 0, 1, \ldots, r$, $\sum_0^1 Q_j = 1$. We then have the following proposition.

Proposition 1:

In the related problem

$$P\{\text{graph is connected}\} = Q_0 .$$

Proof:

The proof is by induction on r and as it is obvious for $r = 1$ assume the result for all values less than r. Now in the case under consideration condition on the set of random variables $Y(i)$ which equal 0 to obtain

$$P\{\text{graph is connected}\}$$

$$= \sum_{S \subset \{1,2,\ldots,r\}} Q_0^{|S|} (1 - Q_0)^{|S^c|} P\{\text{connected} \mid Y(i) = 0, i \in S, Y(i) \neq 0, i \in S^c\}$$

where $|S|$ denotes the cardinality of S and S^c the complement of S. Now given that $Y(i) = 0$ for $i \in S$ and $Y_i \neq 0$ for $i \in S^c$ the situation (as far as the graph being connected) is the same as if we had $|S^c| + 1$ nodes and $|S^c|$ arcs with each arc going into node 0 with probability $\sum_{i \in S} Q_i / (1 - Q_0)$. Hence by the induction hypothesis we have

$$P\{\text{connected} \mid Y(i) = 0 , i \in S , Y_i \neq 0 , i \in S^c\}$$

$$= \sum_{i\in S} Q_i/(1 - Q_0)$$

and so

$$P\{\text{graph is connected}\} = \frac{\sum_S \left[Q_0^{|S|}(1-Q_0)^{|S^c|} \sum_{i\in S} Q_i\right]}{1 - Q_0}$$

$$= \frac{E\left[\sum_{i:Y(i)=0} Q_i\right]}{1 - Q_0}$$

$$= \frac{E\left[\sum_{i=1}^{r} Q_i I_i\right]}{1 - Q_0} \quad \text{where } I_i = \begin{cases} 1 & \text{if } Y_i = 0 \\ 0 & \text{if } Y_i \neq 0 \end{cases}$$

$$= \frac{\sum_1^r Q_i Q_0}{1 - Q_0}$$

$$= Q_0 \;.||$$

Consider now the original problem with n nodes and arcs $(i,X(i))$, $i = 1,2, \ldots, n$. Starting at some node - say node 1 - consider the sequence of nodes $1,X(1),X^2(1), \ldots$ where $X^n(1) = X(X^{n-1}(1))$; and define N by

$$N = \text{smallest } k : X^k(1) \in \{1,X(1), \ldots, X^{k-1}(1)\}$$

and define W by

$$W = P_1 + \sum_{k=1}^{N-1} P_{X^k(1)} .$$

In other words N is the number of nodes reached in the sequence $1, X(1), X^2(1), \ldots$ before a node appears twice and W is the sum of the probabilities of these nodes. We now have

<u>Theorem 1</u>:

$$P\{\text{graph is connected} \mid W\} = W .$$

<u>Proof</u>:

Conditioning on W and N the problem reduces to the related problem and the result follows from Proposition 1.

Hence we have

<u>Corollary 1</u>:

$$P\{\text{graph is connected}\} = E(W) .$$

In other words if a sequence of independent trials each resulting in one of n possible outcomes with probabilities $P_1, \ldots, P_n$ are performed then, given that the initial outcome is outcome 1, the expected sum of the probabilities of all the distinct outcomes obtained before any outcome has been repeated twice is equal to the probability of the graph being connected. It is interesting to note that as we could have begun with any of the n outcomes it follows that the expected sum obtained is independent of the initial outcome; a result which is not all apparent. Hence if we assume that the initial outcome is also randomly determined we have that

Corollary 2:

$$P\{\text{graph is connected}\}$$

$$= \sum_i P_i^2 \left[1 + \sum_{j\neq i} P_j + \sum_{j\neq i} \sum_{k\neq i,j} P_j P_k + \sum_{j\neq i} \sum_{k\neq i,j} \sum_{\ell\neq i,j,k} P_j P_k P_\ell + \dots\right] .$$

Proof:

The term inside the sum is just P_i multiplied by the probability of a type i outcome before any of the other outcomes have occurred twice.||

A graph is said to consist of r connected components if its nodes can be partitioned into r subsets so that each of the subsets are connected and there are no arcs between nodes in different subsets. Let C denote the number of connected components of the random graph $(i,X(i))$, $i = 1, \dots, n$; and let

$$f_j(\underline{P}) = P\{C = j\} \text{ , } j = 1,2, \dots, n$$

where we use the notation $f_j(\underline{P})$ to make explicit the dependence on the probability vector $\underline{P} = (P_1, \dots, P_n)$. Now

$$f_1(\underline{P}) = P\{C = 1\} = P\{\text{graph is connected}\}$$

can be obtained from Corollary 2. To determine $f_2(\underline{P})$, the probability of exactly 2 components fix attention on some particular node - say node 1. In order that a given set of nodes containing node 1 - call it S - will constitute one connected component and the remaining nodes S^c a second connected component we must have

(i) $X(i) \in S \; , \; i \in S$

(ii) $X(i) \in S^c \; , \; i \in S^c$

(iii) The nodes in S form a connected subgraph

(iv) The nodes in S^c form a connected subgraph.

Hence the probability of 2 connected components is given by

$$(1) \qquad f_2(\underline{P}) = \sum_S \left[\left(\sum_{i \in S} P_i \right)^{|S|} \left(\sum_{i \notin S} P_i \right)^{n-|S|} f_1(\underline{P}(S)) f_1(\underline{P}(S^c)) \right]$$

where the sum is over all the subsets S containing node 1 and the i^{th} component of $\underline{P}(S)$ is equal to $P_i / \sum_{j \in S} P_j$ if $i \in S$ and is 0 if $i \notin S$ and similarly for $\underline{P}(S^c)$. In general the recursive formula for $f_j(\underline{P})$ is given by

$$(2) \qquad f_j(\underline{P}) = \sum_S \left[\left(\sum_{i \in S} P_i \right)^{|S|} \left(\sum_{i \notin S} P_i \right)^{n-|S|} f_1(\underline{P}(S)) f_{j-1}(\underline{P}(S^c)) \right]$$

where again the sum is over all subsets of $\{1,2, \ldots, n\}$ which contain node 1.

The expected number of connected components can most easily be computed by first noting that every connected component will contain exactly one cycle. This is most easily seen by noting that a connected component having r nodes will also have r arcs and thus exactly one cycle. Hence

$$E(C) = E(\text{Number of Cycles})$$

$$(3) \qquad = E\left[\sum_S I(S) \right] \quad \text{where } I(S) = \begin{cases} 1 & \text{if nodes in } S \\ & \text{constitute a cycle} \\ 0 & \text{otherwise} \end{cases}$$

$$= \sum_S (|S| - 1)! \left(\prod_{j \in S} P_j \right)$$

where the sum is over all nonempty subsets of $\{1,2, \ldots, n\}$.

The variance of C can also be obtained by using the same representation. We obtain

$$\begin{aligned} \operatorname{Var}(C) &= \operatorname{Var}\left[\sum_S I(S)\right] \\ &= \sum_S \operatorname{Var}[I(S)] + \sum_{\substack{S,S' \\ S' \neq S}}\sum \operatorname{Cov}[I(S), I(S')] . \end{aligned}$$

Now for $S \neq S'$,

$$E[I(S)I(S')] = \begin{cases} 0 & \text{if } S \cap S' \neq \emptyset \\ E[I(S)]E[I(S')] & \text{if } S \cap S' = \emptyset \end{cases}$$

and thus

$$\text{(4)} \quad \operatorname{Var}(C) = \sum_S E[I(S)](1 - E[I(S)]) - \sum_S E[I(S)] \sum_{\substack{S' \neq S \\ S' \cap S \neq \emptyset}} E[I(S')] .$$

Now,

$$\sum_{S' \cap S \neq \emptyset} I(S') = \text{Number of cycles having a node in } S$$

and thus

$$E\left[\sum_{\substack{S' \cap S \neq \emptyset \\ S' \neq S}} I(S')\right] = E(C) - E[I(S)] - E\left[\sum_{S' \subset S^c} I(S')\right] .$$

Hence from (4) we obtain

$$(5) \quad \mathrm{Var}\ (C) = \sum_{S} E[I(S)] - \left[\sum_{S} E[I(S)]\right]^2 + \sum_{S,S' \subset S^c} \sum E[I(S)]E[I(S')]$$

where

$$E[I(S)] = (|S| - 1)! \prod_{j \in S} P_j \ .$$

2. THE SPECIAL CASE $P_j \equiv 1/n$

In the special case where $P_j = 1/n$ for all j we have from Corollary 2 that

$$P\{\text{graph is connected}\}$$

$$= \frac{1}{n}\left[1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots + \frac{(n-1)!}{n^{n-1}}\right]$$

$$= \frac{1 + \sum_{j=1}^{n-1}\left[\frac{(n-1)!}{(n-j-1)!\,n^j}\right]}{n}$$

$$= \frac{(n-1)!}{n^n}\sum_{i=0}^{n-1} n^i/i! \qquad \text{(by letting } i = n - j\text{)}$$

which agrees with the results given in [1], [3] and [4]. In addition letting $P_n(j) = f_j\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = P\{C = j\}$ the recursive formulae (1) and (2) become

$$P_n(2) = \sum_{k=1}^{n-1}\binom{n-1}{k-1}\left(\frac{k}{n}\right)^k\left(\frac{n-k}{n}\right)^{n-k} P_k(1)P_{n-k}(1)$$

and, in general,

$$P_n(j) = \sum_{k=1}^{n-j+1}\binom{n-1}{k-1}\left(\frac{k}{n}\right)^k\left(\frac{n-k}{n}\right)^{n-k} P_k(1)P_{n-k}(j-1)\ .$$

Whereas an explicit expression for $P_n(j)$ in terms of Stirling numbers has been previously derived in (1) the above recursive equations appears to be new. In this special case the formula (3) for $E(C)$ simplifies to

$$E(C) = \sum_{k=1}^{n} \binom{n}{k}(k - 1)!/n^k$$

which had previously been obtained (see [1] or [5]) in a much more involved manner. In addition, from (5) we have

$$\mathrm{Var}\ (C) = \sum_{k=1}^{n} \binom{n}{k}(k - 1)!/n^k - \left(\sum_{k=1}^{n} \binom{n}{k}(k - 1)!/n^k\right)^2$$
$$+ \sum_{k=1}^{n} \binom{n}{k} \frac{(k - 1)!}{n^k} \sum_{j=1}^{n-k} \binom{n - k}{j} \frac{(j - 1)!}{n^j}$$

which appears to be new.

3. SCHUR FUNCTIONS

We say that the vector $\underline{P} = (P_1, \ldots, P_n)$ majorizes the vector $\underline{Q} = (Q_1, \ldots, Q_n)$, written $\underline{P} \underset{m}{\geq} \underline{Q}$ if

$$\sum_{j=1}^{i} P_{(j)} \geq \sum_{j=1}^{i} Q_{(j)} \text{ , } i = 1, \ldots, n-1$$

$$\sum_{j=1}^{n} P_{(j)} = \sum_{j=1}^{n} Q_{(j)}$$

where $P_{(j)}$ and $Q_{(j)}$ are respectively the j^{th} largest values of $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$.

A permutation invariant function $H(\underline{P})$ is said to be a Schur convex function if $H(\underline{P}) \geq H(\underline{Q})$ whenever $\underline{P} \underset{m}{\geq} Q$. It is well known (see [2]) that a sufficient condition for H to be Schur convex is that

$$(P_1 - P_2)\left[\frac{\partial}{\partial P_1} H(\underline{P}) - \frac{\partial}{\partial P_2} H(\underline{P})\right] \geq 0$$

for all P_1, P_2.

It is straightforward to verify that the expression presented in Corollary 2 for the probability that the graph is connected is thus Schur convex and from Equation (3) that $-E(C)$ is Schur convex. Since every probability vector $\underline{P} = (P_1, \ldots, P_n)$ majorizes the vector $\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ it follows that the probability that the graph is connected is minimized and the expected number of components is maximized by the results in the uniform case.

REFERENCES

[1] Frank, H. and I. Frisch, COMMUNICATION, TRANSMISSION AND TRANSPORTATION NETWORKS, Addison-Wesley, 1971.

[2] Hardy, G. H., J. E. Littlewood and G. Polye, INEQUALITIES, Cambridge University Press, Cambridge, MA, 1952.

[3] Harris, B., "Probability Distributions Related to Random Mappings," Annals of Mathematical Statistics, Vol. 31, pp. 1045-1062, (1960).

[4] Katz, I., "Probability of Indecomposability of a Random Mapping Function," Annals of Mathematical Statistics, Vol. 26, pp. 512-517, (1955).

[5] Kruskal, M. D., "The Expected Number of Components Under a Random Mapping Function," American Mathematical Monthly, Vol. 61, pp. 392-397, (1954).